Studies of Scale and Ambiguity Values
Obtained by the Method of Equal
Appearing Intervals

By

Sam C. Webb
Emory University

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Studies of Scale and Ambiguity Values Obtained by the Method of Equal-Appearing Intervals[1]

Sam C. Webb
Emory University

I. INTRODUCTION

THOUGH the method of equal-appearing intervals has been widely used in constructing attitude scales, a recent review of the literature (12) suggests there still remain aspects of it which have not been systematically evaluated. This monograph presents a series of studies which provide additional data concerning some aspects of the method. Specifically, data are provided concerning the following six topics:

1. Efficiency and reliability of estimators of scale and ambiguity values.

2. Interactive effects of number of judges, number of intervals, and efficiency of estimating formulae on the reliability of scale and ambiguity values.

3. Intercorrelations for scale and ambiguity values.

4. Effects of number of judges, number of intervals, and efficiency of estimating formulae on the relation between scale and ambiguity values.

5. Relation of scale and ambiguity values obtained by the procedures of this study to the values obtained by Thurstone and Chave.

6. Relevance of scale and ambiguity values for selecting items for Guttman-type scales.

II. ASSEMBLING THE DATA

The general intent of most of these studies was to investigate the effect of three factors—efficiency of estimating formulae, number of judges, and number of judging intervals—upon various aspects of Thurstone-type scale and ambiguity values. For convenience these factors are hereafter referred to as formulae, judges, and intervals.

The first step was that of deciding how these factors should be varied. A consideration of previous studies suggested that three different numbers of intervals, three different numbers of judges, and four different formulae for computing scale and ambiguity values should be employed. The number of intervals selected were 11, 7, and 5. The number of judges selected were 25, 50, and 100. The computing formulae were the mean and three percentile estimates of the mean for scale values, and the standard deviation and three percentile estimates of the standard deviation for ambiguity values. These computing formulae are described in more detail in the next section.

The second step was that of selecting a set of attitude items. For this study the 130 statements of attitude toward the church of Thurstone and Chave (10) seemed appropriate. This set had the advantages of

being relatively large in number and of having adequate representation in all intervals of the scale.[2]

The third step was that of assimilating a pool of judgments of these items from which scale and ambiguity values based on various judges-intervals-formulae combinations could be obtained.

A. Collection of Judgment Data

The method of collecting the judgments of the items was a modification of that used by Thurstone and Chave. Judgments using 11, 7, and 5 intervals were recorded on specially prepared IBM answer sheets which provided appropriate numbers of identified intervals. The original Thurstone and Chave directions (10), modified to provide for recording ratings on the answer sheet, were used. Ratings were made by groups of 20 to 30 students during a 50-minute class period. The directions were read aloud to the students, and all questions concerning them were answered before the rating of the items started. Special precautions which were taken to insure a minimum of error associated with a failure to follow directions have been described elsewhere (13).

B. Subjects

The 11-interval data were collected from 243 students enrolled in psychology classes at the Georgia Institute of Technology. The 7-interval data were collected from 244 students enrolled in psychology classes at the University of Georgia; and the 5-interval data were collected from 225 students enrolled in psychology and sociology classes at Emory University.

C. Selection of Subsamples

All sortings were next examined for ir-

regularities; on the basis of a classification of irregularities (13) sortings with the most serious irregularities were eliminated until the number of sortings remaining in each set of data was reduced to 200.

Sortings in each group were then numbered 1 to 200. Finally for each set of data, samples of 25, 50, and 100 sortings were selected according to the following procedure: Using a table of random numbers, 25 sortings were drawn from the pool of 200 sortings for the 11-interval data. This subsample was numbered 11-I-25 to indicate it was composed of 11-interval sortings and that it was sample number 1 of size 25. Without replacing these sortings, a second set of 25 sortings was drawn by using the table of random numbers and labeled 11-II-25 to indicate it was the second sample of size 25. These two samples were combined to form sample 11-I-50, the first sample of size 50; and a second sample of size 50, labeled 11-II-50, consisted of 50 sortings selected at random from the remaining 150 sortings in the pool. Finally the two samples of size 50 were combined to form a sample labeled 11-I-100. The remaining 100 sortings formed sample 11-II-100. Similar procedures were followed in selecting samples for the 7-interval and 5-interval data. The procedure therefore yielded a total of 18 samples, two for each of the nine interval-judge combinations. For each sample the frequency with which each item fell in each category was determined. Cumulative proportions were determined, ogives drawn, and scale and ambiguity values computed by procedures described in the following section.[3]

III. Efficiency and Reliability of Estimators of Scale and Ambiguity Values

In developing the method of equal-ap-

---

[2] While Edwards (1) has demonstrated the ineffectiveness of neutral items, these items were included to prevent gaps in the distributions of indices of scale value.

[3] Items left unrated by subjects were arbitrarily placed in the neutral category.

pearing intervals Thurstone assumed that the distribution of judgments for each item when plotted on the attitude scale is described by the phi-gamma function (9). While this assumption implies that the mean and standard deviation should be the most reliable estimates of scale and ambiguity value, the median and interquartile range have generally been used for computational purposes. Except for a comparison of the mean and median as estimates of scale value, little attention has been directed toward the problem of determining the need for, or possibility of, using other statistics as indices of scale and ambiguity values.

A. Purpose

On the basis of Thurstone's assumption of the distribution of judgments for each item plus the demonstrations by Kelley (7), Mosteller (8), and Yost (15) that the efficiency of various combinations of percentile values for estimating the mean and standard deviation of a normal supply increases as the properly selected number of percentile points included in the estimating formula increases, it should follow that the mean and standard deviation should be the most reliable estimates of scale and ambiguity value and that estimates of these based on successively less efficient estimating formulae should show successively less reliability. The purpose of this study therefore was to investigate to what extent the use of successively more efficient estimating formulae would be accompanied by successively higher reliability of the estimates. Attention is focused primarily on the ordering of the estimates of reliability according to magnitude. The problem of significance of differences is reserved for a later section.

B. Choice of Formulae and Procedures

For measures of scale value the mean and three estimating formulae based on

percentile values were used. These formulae and the identifying symbols for them used in this study were $M_1 = P_{50}$, $M_2 = .5(P_{25} + P_{75})$, and $M_3 = .333(P_{20} + P_{50} + P_{80})$. These have an efficiency of .64, .81, and .88 respectively (8). For measures of ambiguity the standard deviation and three estimating formulae based on percentile values were used. These formulae and the identifying symbols for them used in this study were $S_1 = .5(P_{75} - P_{25})$, $S_2 = .3388(P_{93} - P_{07})$, and $S_3 = .2157(P_{93} + P_{80} - P_{20} - P_{07})$. These have an efficiency of .37, .65, and .75 respectively (8).

For each of the two samples containing 100 judges for the 11-, 7-, and 5-interval data, percentile values required for the three estimating formulae were obtained by dropping perpendiculars at the specified points. It should be noted that Thurstone and Chave's procedure of extrapolating for P_{50} and of doubling the quartile distance for estimating ambiguity values for extreme items was not followed.

While the general hypotheses to be tested in this study were that the ordering of reliability of estimates of scale value should be $M > M_3 > M_2 > M_1$ and that the ordering of reliability of estimates of ambiguity should be $S > S_3 > S_2 > S_1$, it was necessary to state these hypotheses slightly differently for the two procedures used in treating the data.

According to the first procedure the three percentile estimates of scale value were correlated with the item means and the three percentile estimates of ambiguity were correlated with the item standard deviations for each sample. These results are shown in Table 1. In terms of these data each successively more efficient percentile estimate of scale and ambiguity value should provide successively higher correlations with the mean and standard deviation respectively. Accordingly hypotheses could be stated as follows: *Hypothesis I.* The

TABLE 1

CORRELATIONS OF PERCENTILE ESTIMATES OF SCALE AND AMBIGUITY
VALUE WITH MEANS AND STANDARD DEVIATIONS

	Sample I—100				Sample II—100			
	M vs.—		S vs.—		M vs.—		S vs.—	
11 Interval	M_1	.9979	S_1	.732	M_1	.9968	S_1	.760
	M_2	.9991	S_2	.907	M_2	.9989	S_2	.928
	M_3	.9994	S_3	.893	M_3	.9992	S_3	.922
7 Interval	M_1	.9954	S_1	.757	M_1	.9973	S_1	.722
	M_2	.99856	S_2	.904	M_2	.9988	S_2	.891
	M_3	.99859	S_3	.885	M_3	.9990	S_3	.862
5 Interval	M_1	.9965	S_1	.758	M_1	.9949	S_1	.676
	M_2	.9978	S_2	.864	M_2	.9970	S_2	.884
	M_3	.9982	S_3	.885	M_3	.9976	S_3	.853

ordering of the correlations of M_1, M_2, and M_3 with M is r_M vs. $M_3 > r_M$ vs. $M_2 > r_M$ vs. M_1. *Hypothesis II*. The ordering of the correlations of S_1, S_2, and S_3 with S is r_S vs. $S_3 > r_S$ vs. $S_2 > r_S$ vs. S_1.

In the second procedure values for the mean, standard deviation, and percentile estimates of these for the two samples of each interval size were correlated so as to provide reliability coefficients for each estimate of scale and ambiguity value. These data are shown in Table 2. On the basis of this procedure two hypotheses could be stated as follows: *Hypothesis III*. The ordering of the reliability coefficients for estimates of scale value should be $M > M_3 > M_2 > M_1$. *Hypothesis IV*. The ordering of the reliability coefficients for

TABLE 2

RELIABILITY COEFFICIENTS FOR ESTIMATORS
OF SCALE AND AMBIGUITY VALUE

	11 Interval	7 Interval	5 Interval
M	.9962	.9951	.9925
M_3	.9964	.9952	.9920
M_2	.9959	.9949	.9913
M_1	.9956	.9950	.9906
S	.800	.711	.749
S_3	.804	.860	.778
S_2	.870	.804	.721
S_1	.841	.807	.710

estimates of ambiguity value should be $S > S_3 > S_2 > S_1$.

In view of the interest in the literature in the comparability of median vs. mean and standard deviation vs. interquartile range as measures of scale and ambiguity value, two additional hypotheses could be tested on the basis of these data. *Hypothesis V*. There is no difference in reliability between the mean and median as estimates of scale value. *Hypothesis VI*. There is no difference in reliability between the standard deviation and the interquartile range as estimators of ambiguity value.

C. Results

The data of Table 1 show that for all six samples the ordering of the correlations of the percentile estimates with the mean is in the predicted order. Hypothesis I is therefore confirmed. Since the ordering of the correlations of percentile estimates with the standard deviation follows the predicted order in only one of the six samples, Hypothesis II is *not* confirmed.

The data in Table 2 show that the estimates of scale value fall in the predicted order only once in three trials. Since the probability of obtaining the predicted ordering at least once in three replications

by chance is .115, Hypothesis III is *rejected*. Since the specified order of reliability coefficients for the estimates of ambiguity does not appear at all, Hypothesis IV is *not* sustained. To test Hypothesis V the coefficients for M and M_1 of Table 2 were converted to Fisher's z. On the basis of this transformation an analysis of variance was performed which yielded an F of 4.187. Since a one-tailed test is appropriate for testing this hypothesis, the F was converted to $t = 2.04$ which for $df = 2$ is significant at the 5% level of confidence. Since S is predicted to be greater than S_1 and since it was greater only one in three trials, Hypothesis VI also is *not* confirmed.

D. Discussion

Despite the fact that only two of the three hypotheses relative to estimates of scale value were confirmed, it seems safe to assert that the data show a trend for scale values computed from successively more efficient formulae to show a successively higher reliability. However, since even the least efficient estimate M_1 provides reliability coefficients greater than .99, the increased reliability of scale values resulting from the use of more efficient formulae is of little practical consequence.

While none of the hypotheses concerning the ambiguity values were sustained, the data suggest that two of the computing formulae—S_2 and S_3—yield more reliable estimates of ambiguity than are afforded by the commonly used interquartile range. It is not clear, however, to what extent the violation of the assumption of normality of distribution of judgments per item, caused by a lack of correction for the influence of the end effect, has prevented the reliabilities of these formulae from being higher than the obtained values. It is possible, as Kelley (7) has suggested, that if some other form of distribution which more nearly represents the modal-type distribution of judgments for attitude items were used as the basis of deriving estimating formulae, it might be possible to obtain higher reliabilities than those obtained by the formulae used here.

An unexpected finding about ambiguity values was that in two out of three sets of data the reliability of S was *lower* than that of S_1. Inspection of the scattergrams shows that a partial cause of these low coefficients is a wide discrepancy of the two S values for a few items. In every case these items proved to be extreme items which a few judges of one sample but not the other had judged to fall at one extreme of the continuum, while all other judges had judged them to fall at the other end. It appears, therefore, that the great sensitivity of S to extreme deviates makes it unsuitable as an estimate of ambiguity for data collected by the method of equal-appearing intervals.

IV. EFFECTS OF NUMBER OF JUDGES, NUMBER OF SCALE INTERVALS, AND EFFICIENCY OF ESTIMATING FORMULAE ON THE RELIABILITY OF SCALE AND AMBIGUITY VALUES

The purpose of this study was to investigate within the framework of a three-dimensional factorial design the main and interaction effects of three variables on the reliability of scale and ambiguity values. The variables selected were *number of judges, number of scale intervals*, and *estimating formulae*.

On the basis of previous research reports and statistical theory it was expected that scale and ambiguity values would show increased reliability as a function of increasing number of judges, increasing number of scale intervals, and increasing efficiency of estimating formulae. As has previously been stated 11-, 7-, and 5-interval scales and groups of 25, 50, and 100 judges were used. On the basis of the results of the preceding study, three estimating formulae for

scale values—M_1, M_2, and M_3—and three estimating formulae for ambiguity values—S_1, S_2, and S_3—were used. These formulae have been described in section III.

A. Design and Procedure

For each of the two samples for each of the nine judge-interval combinations, three estimates of the scale and ambiguity values were computed by procedures already described. The reliability of the various estimates of scale and ambiguity values for each judge-interval-formula combination was then determined by computing the Pearson product-moment correlation between the appropriate values for the 130 items of the two samples for each judge-interval-formula combination. The resulting coefficients for scale and ambiguity values are shown in Tables 3 and 4, respectively. Finally the coefficients of each table were transformed into Fisher's z, and an analysis of variance was performed on these transformed values. The results of these analyses are shown in Tables 5 and 6.

B. Results

Scale Values. All the reliability coefficients for scale values fell within the narrow range of .9760 and .9964. The consistency of ordering of these coefficients across columns in the direction of increased reliability with increase in efficiency of the computing formula, across rows in the direction of increased reliability with increase in number of judges, and across blocks in the direction of increased reliability with increasing number of intervals, should be noted.

When tested against the triple interaction term as error, only one of the three first-order interactions (judges by intervals) is significant. This one interaction is significant at a 1% level of confidence. As for the main effects, when the variance for columns (formulae) is tested against the triple interaction term as error, the null

TABLE 3

RELIABILITY OF SCALE VALUES FOR THE INTERVAL-JUDGE-FORMULA COMBINATIONS

No. Inter-vals	No. Judges	Estimating Formulae		
		M_1	M_2	M_3
11	25	.9866	.9856	.9858
	50	.9927	.9941	.9942
	100	.9956	.9959	.9964
7	25	.9804	.9819	.9820
	50	.9896	.9918	.9924
	100	.9950	.9949	.9952
5	25	.9760	.9764	.9779
	50	.9825	.9814	.9832
	100	.9906	.9913	.9920

hypothesis of no difference among the column means is rejected at a 5% level of confidence. Though the null hypothesis of no difference among means for judges and no difference among means for intervals cannot, in the strictest sense, be made because of the significant judges-by-interval interaction, it nevertheless seems fair to assert that the main effects of judges and intervals are significant sources of variance. This assertion is based on the fact that highly significant F ratios are obtained by dividing the judges and intervals variances by the variance of the appropriate interaction terms. These high ratios are no doubt a function of the consistency of

TABLE 4

RELIABILITY OF AMBIGUITY VALUES FOR INTERVAL-JUDGE-FORMULA COMBINATIONS

No. Inter-vals	No. Judges	Estimating Formulae		
		S_1	S_2	S_3
11	25	.517	.560	.560
	50	.763	.807	.825
	100	.841	.870	.894
7	25	.571	.512	.607
	50	.688	.686	.762
	100	.807	.804	.860
5	25	.577	.543	.573
	50	.713	.694	.733
	100	.710	.721	.778

TABLE 5

ANALYSIS OF VARIANCE OF FISHER'S z TRANSFORMATION OF SCALE
VALUE RELIABILITY COEFFICIENTS

Source	Est.	df	ΣX^2	Variance	F_{rbc}*	F_{bc}	F_{rb}	F_{rc}
Rows (Judges)	S_r^2	2	1.592551	.796275			47.613 (.001)	460.757 (.001)
Blocks (Intervals)	S_b^2	2	.691235	.345618		888.478 (.001)	20.666 (.01)	
Columns (Formulae)	S_c^2	2	.016861	.008431	6.227 (.05)			
$R \times B$	S_{rb}^2	4	.066895	.016724	12.352 (.01)			
$R \times C$	S_{rc}^2	4	.005678	.001420	1.049 (—)			
$B \times C$	S_{bc}^2	4	.001555	.000389	.287 (—)			
$R \times B \times C$	S_{rbc}^2	8	.010833	.001354				
Total		26	2.385608					

* Subscript indicates interaction used as error variance in F ratio.

ordering across rows, columns, and blocks noted above.

Finally, t tests between the z values for the various pairs of each of the three main effects were computed. Using judges-by-intervals interaction as error, the differences between all possible pairings of the three levels for judges were found to be sig-

TABLE 6

ANALYSIS OF VARIANCE OF FISHER'S z TRANSFORMATION OF AMBIGUITY
VALUE RELIABILITY COEFFICIENTS

Source	Est.	df	ΣX^2	Variance	F_{rbc}*	F_{rc}	F_{bc}	F_{rb}
Rows (Judges)	S_r^2	2	1.249510	.624755		154.834 (.001)		16.602 (.05)
Blocks (Intervals)	S_b^2	2	.160822	.080411			16.085 (.05)	2.137 (—)
Columns (Formulae)	S_c^2	2	.069218	.034609		8.577 (.05)	6.923 (.05)	
$R \times B$	S_{rb}^2	4	.150528	.037632	163.617 (.001)			
$R \times C$	S_{rc}^2	4	.016140	.004035	17.543 (.001)			
$B \times C$	S_{bc}^2	4	.019997	.004999	21.735 (.001)			
$R \times B \times C$	S_{rbc}^2	8	.001842	.000230				
Total		26	1.668056					

* Subscripts indicate interaction used as error variance in F ratio.

nificant at a .001 level of confidence (two-tailed test). These differences were all in the expected directions.

Again using this same interaction as error, the differences between all possible pairings of the three levels for intervals were tested and found significant at a 5% or higher level of confidence. These differences were all in the expected directions. Using the second-order interaction as error, differences between all possible pairings of formulae were found to be significant at a 1% level of confidence in the expected directions.

Ambiguity Values. The reliability coefficients for ambiguity values ranged from .517 to .894. In Table 4 a high consistency of ordering is noted only across rows, where increased reliability with increased number of judges is found. Especially among the coefficients based on 25 judges there appears to be little systematic variation.

Using the second-order interaction as error, all first-order interactions are significant at a .001 level of confidence. Though no tests of the null hypothesis of no differences among means for judges, intervals, or formulae can, in the strictest sense, be made because of the significant interactions, the fact that all but one of the ratios obtained by dividing the variances for main effects by the appropriate first-order interactions are fairly large (significant at a 5% level of confidence) suggests that the main effects are significant sources of variance over and above the variance contributed by the interactive effects.[4] For this reason the significance of difference among main effects was further investigated by the use of t tests. When both judges-by-formulae and intervals-by-formulae interactions are used as error the

[4] The F ratio for formulae over interval-by-formulae interaction actually falls below the 5% value by .017. Because of this slight difference it has for practical purposes been called significant at a 5% level of confidence.

differences between the average z values for the three formulae were all in the expected directions; S_2 was not found significantly different from S_1 (two-tailed test), but S_3 was found significantly different from S_1 and S_2.

Using judges-by-intervals and judges-by-formulae variance as error, all differences between judges were in the expected direction and all were significant at a 5% or higher level of confidence. All differences between number of intervals were in the expected directions; but using judges-by-intervals interaction as error the average z for 11 intervals was significantly different from average z for 5 intervals at a 5% or higher level. Only when intervals-by-formulae interaction was used as error were all differences significant at a 5% or higher level of confidence.

C. Discussion

From the data of Tables 5 and 6 it seems clear that the rank order of judges, intervals, and formulae for contributing to reliability would be the same as their rank order for contributing to the total sum of squares—number of judges first, number of intervals second, and formulae third. A possible cause for the low ranking of formulae lies in the design of the experiment. For since all the estimates computed from the formulae were based on the ogives for the nine judge-intervals combinations, the requirement of randomness is not fully satisfied; and the three estimates of scale and of ambiguity values are to some extent correlated. This is especially so for M_1 and M_3, since they have one percentile value in common, and for S_2 and S_3, since they have two percentile values in common. It should be emphasized, however, that even when no adjustment is made for this correlation effect, the F ratios suggest significant differences; and the t tests between the most correlated estimates are significant.

It should be further noted that schools (sources of the data) have been confounded with intervals; but since there is no obvious reason to question the comparability of the three groups of students in regard to ability to judge the items, this is not a serious defect of the design.

In the planning of this study the hope was entertained that the reliabilities for the various judge-interval-formula combinations would be such that combinations involving more efficient formulae and fewer judges and/or intervals would produce reliabilities on a par with those obtained by using less efficient formulae and more judges and/or intervals. Such findings would suggest alternate methods of obtaining scale and ambiguity values of a desired level of reliability with a reduction of tedium involved in collecting and processing data. Considered from this practical viewpoint, estimates of scale value computed by the formula M_1 on the basis of the judgments of 25 judges using a 5-interval scale appear to be as satisfactory as those computed by more efficient formulae on the sortings of more judges using more intervals. These results are in accord with the finding of other investigations (13, 14).

While the results for ambiguity values are not exactly what was hoped for, depending on the level of reliability desired, some choice of procedure seems evident. For example, if an investigator requires a reliability of .55 to .60 for his ambiguity values, estimates computed by formula S_1 on the basis of the judgments of 25 judges using 5 intervals would be satisfactory. Values with a reliability of approximately .75 could be obtained using 50 judges, 11 intervals, and formula S_1; or 50 judges, 7 intervals, and formula S_3; or perhaps even 100 judges, 5 intervals, and formula S_3. Reliabilities of approximately .80 could be obtained using 50 judges, 11 intervals, and

formula S_2; or 100 judges, 7 intervals, and formula S_1.

V. INTERCORRELATIONS FOR SCALE AND AMBIGUITY VALUES

Before recommending the use of any judge-interval-formula combination which differs from the conventionally used combination of 11 intervals, approximately 100 judges, and M_1 and S_1 formulae, a comparison should be made of the values obtained by the conventional and possible alternate combinations concerning similarity of ordering of items, in addition to reliability. Data for such a comparison are provided by intercorrelations among values computed on the basis of various judge-interval-formula combinations. On the basis of the work of other investigators and data previously presented in this study, it would be expected that all intercorrelations among scale values would be high, while the intercorrelations among ambiguity values would be more variable. The purpose of this section was to determine the extent to which these expectations are confirmed by experimental data.

A. Procedure

Rather than compute all possible intercorrelations, only those necessary to demonstrate these expectations were computed. In general, the principle was followed of correlating values of such judge-interval-formula combinations as seemed appropriate with the values obtained by the combination conventionally employed and with values obtained by combinations which yield higher reliability than that of the conventionally used combinations. Accordingly, scale values computed by formulae M, M_1, M_2, and M_3 from sample 11-II-100 were correlated with the scale values obtained by formulae M_1, M_2, and M_3, for sample 5-II-25. These are shown in the top half of Table 7. In addition, scale values

TABLE 7

INTERCORRELATION BETWEEN SCALE VALUES OF SELECTED SAMPLES AND SCALE VALUES OF SAMPLE 5-II-25

Sample		Sample 5-II-25		
		M_3	M_2	M_1
11-II-100	M	.9807	.9790	.9773
	M_3	.9820	.9804	.9793
	M_2	.9813	.9796	.9785
	M_1	.9823	.9810	.9817
11- I-100	M	.9959	.9954	.9948
7- I-100	M	.9951	.9951	.9927
5- I-100	M	.9909	.9903	.9889

computed by the formula M on samples 11-I-100, 7-I-100, and 5-I-100 were correlated with scale values computed by formulae M_1, M_2, and M_3 for sample 5-II-25. These are shown in the bottom half of Table 7. For ambiguity values estimates computed by formulae S_1, S_2, and S_3 on sample 11-II-100 were correlated with each other and with values computed by formulae S_1, S_2, and S_3 for every interval-judge combination for samples numbered II. In addition, values computed by formulae S_1, S_2, and S_3, on sample 7-II-100 were correlated with each other; and values similarly computed on sample 5-II-100 were correlated with each other. All these intercorrelations are shown in Table 8.

B. Results

Scale Values. Estimates computed from sample 11-II-100 and the mean values computed on samples 11-I-100, 7-I-100, and 5-I-100, correlate approximately .98 and .99, respectively, with all values computed on the basis of sample 5-II-25.

Ambiguity Values. The intercorrelations for ambiguity values show considerable variability. These range from .98 to .57. But except where indices are correlated with other indices computed from the same judge-interval combination, no intercorrelation is higher than .83. With few exceptions S_1 correlates higher with S_3 than with S_2; and S_3 correlates higher with S_2 than S_1. In those instances where values computed on the 11-interval 100-judge combination are correlated with values computed on other judge-interval combinations there is a general tendency for: (a) correlations between S_3 values to be highest, (b) correlations between S_2 values to be higher than correlations between S_1 values, (c) correlations with values computed on sortings of 50 judges to be lower than correlations with values based on 100 judges, and (d) correlations with values computed on sortings of 25 judges to be lower than correlations with values based on 50 judges.

C. Discussion and Summary

For scale values the intercorrelation coefficients seem high enough to indicate that values computed by the formulae M_1, M_2, or M_3 on sortings of 25 persons using 5

TABLE 8

INTERCORRELATIONS BETWEEN AMBIGUITY VALUES FOR SELECTED INTERVAL-JUDGE-FORMULAE COMBINATIONS

		11									7									5								
		100		50			25			100			50			25			100			50			25		
		S_3	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1
11 100	S_1	98	88	85	81	78	70	67	62	83	80	79	76	71	75	75	68	66	77	76	68	80	80	66	71	71	64
	S_3		77	81	81	71	69	66	58	80	80	72	72	68	66	75	68	64	74	74	64	78	80	62	68	68	61
	S_2			77	77	69	60	57	64	75	67	80	72	63	76	66	58	60	70	66	66	71	69	64	64	63	59
7 100	S_3									94	89																
	S_2										80																
5 100	S_3																		93	85							
	S_2																			71							

intervals order the items in essentially the same way as do indices computed by these same or more reliable formulae on samples using more judges and more intervals.

For ambiguity values the intercorrelations of value computed on the 11-interval 100-judge sample correlate from .83 to .57 with values computed on the basis of the other interval-judge combinations used in this study. In view of the variability among these correlations, it seems advisable to determine how much departure from the ordering of items obtained from a particular interval-judge-formula combination one will accept before deciding what interval-judge-formula computing combination one might use instead.

VI. RELATION BETWEEN SCALE AND AMBIGUITY VALUES

The purpose of this section was twofold. First, an effort was made to evaluate the effects of sample size, number of intervals, and computing formulae on the magnitude of the relation between scale value and ambiguity values. Second, a comparison was made between the magnitude of the relation between scale and ambiguity values derived from values obtained by the computing methods of this study and the magnitude of the relation for values obtained by the Thurstone and Chave computing methods.

A. Procedure

The fact that the relationship between scale and ambiguity values is curvilinear made it impossible to employ a design permitting an analysis of the variance among the various indices of relationship into parts attributable to the three independent variables about which interest was centered. Instead the less satisfactory procedure was employed of computing the unbiased correlation ratio ϵ of ambiguity values on scale values for a number of interval-judge-formula combinations, and of

attempting to state on the basis of inspection of the data what effect these variables might have on the obtained relationships.

ϵ was computed only for such interval-judge-formula combinations as would be most likely to demonstrate the effects of these variables on the magnitude of the relationship. In view of the high reliability of, and high intercorrelations among, scale values computed by the various formulae, values computed by the M_1 formula were considered as satisfactory measures of scale value. S_1 and S_3 values were considered as sufficient for computing formulae of ambiguity value, and groups of 100 and 25 judges were considered as sufficient for numbers of judges. Accordingly the unbiased correlation ratios of ambiguity value on scale value for samples I and II for selected judge-formula combinations for the 5-, 7-, and 11-interval data were computed. The results are shown in Table 9. Edwards' computation (1) based on the

TABLE 9

UNBIASED CORRELATION RATIOS FOR S_1 AND S_3 AMBIGUITY VALUES ON SCALE VALUES FOR SELECTED INTERVAL-JUDGE COMBINATIONS

In-terval	Sample	M_1 vs.-	K	ϵ^2	ϵ
5	I-100	S_1	14	.475	.689
		S_3	14	.505	.710
	II-100	S_1	14	.514	.717
		S_3	14	.447	.669
	I-25	S_1	15	.243	.493
	II-25	S_1	15	.462	.680
7	I-100	S_1	13	.556	.746
		S_3	13	.521	.722
	II-100	S_1	13	.654	.809
		S_3	13	.591	.769
	I-25	S_1	13	.392	.626
	II-25	S_1	13	.194	.440
11	I-100	S_1	11	.482	.694
		S_3	11	.538	.733
	II-100	S_1	11	.562	.750
		S_3	11	.601	.775
	I-25	S_1	11	.457	.676
	II-25	S_1	11	.178	.422
Thurstone and Chave data		S_1	15	.243	.493

Thurstone and Chave data is also shown in the table.

B. Results

As was anticipated all relationships were significantly curvilinear at a 1% level of confidence with low and high scale values being associated with low ambiguity values.

On the basis of inspection only, the data of Table 9 seem to suggest the following findings: (a) For the 5-, 7-, and 11-interval ϵ values for S_1 ambiguity values on M_1 scale values, the values of ϵ computed on the basis of sortings of 100 judges are higher than the values computed on the basis of the sortings of 25 judges. (b) For samples of 100 judges, ϵ's for S_3 ambiguity values on M_1 scale values do not appear to be systematically higher than values of S_1 ambiguity values on M_1 scale values. (c) The size of ϵ for S_1 ambiguity values on M_1 scale values computed either on the basis of 25 or 100 judges does not appear to be affected systematically by the number of intervals employed in sorting the items. (d) The size of ϵ for S_3 ambiguity values on M_1 scale values computed on the basis of the sortings of 100 judges appears to vary directly with the number of intervals employed in sorting the items. (e) All but two of the 18 ϵ values computed in this study equal or exceed the value obtained by Edwards on the basis of the Thurstone and Chave data.

C. Discussion and Summary

On the basis of these data the following tentative conclusions appear appropriate concerning the effects of number of intervals, number of judges, and computing formulae on the unbiased correlation ratio of ambiguity values on scale values: (a) the size of ϵ for S_1 ambiguity values on M_1 scale values appears to be a function of the number of sortings on which the values are based, but not a function of the number of

intervals employed in sorting the items. (b) For samples of 100 judges the size of ϵ does not appear to be a function of the computing formula used. (c) Size of ϵ for S_3 ambiguity values on M_1 scale values computed on the basis of sortings of 100 judges varies directly as the number of intervals used in sorting the data. (d) The relation of ambiguity value on scale value based on indices computed by the procedures used in this study is higher than that obtained from values obtained by the method used by Thurstone and Chave.

It should be emphasized that because of a lack of any sampling distribution against which to test specific hypotheses, this summary can be regarded as no more than a statement which seems consistent with the empirical data, but which cannot be tested for significance.

VII. RELATION TO SCALE AND AMBIGUITY VALUES OBTAINED BY THURSTONE

It has previously been noted that instead of following the procedure of Thurstone and Chave of extrapolating for P_{50} for estimating scale values and of doubling the quartile distance for estimating ambiguity values for extreme items, the procedure of this study involved dropping perpendiculars at the required percentile points to obtain the values to be used in the various computing formulae. The purpose of this section was to evaluate what effect this procedural difference has on the ordering of items in regard to scale and ambiguity values.

The procedure consisted in computing the following Pearson product-moment correlations: (a) Scale values which Thurstone and Chave obtained on the basis of 300 judges using 11 intervals vs. M_1 scale values obtained on the basis of samples of 100 judges for the 11-, 7-, and 5-interval data of this study (these are shown in column 3 of Table 10). (b) Ambiguity values obtained by Thurstone and Chave vs. the

TABLE 10

CORRELATIONS OF SELECTED SCALE AND AMBI-
GUITY VALUES OF THIS STUDY WITH SCALE
AND AMBIGUITY VALUES OBTAINED BY
THURSTONE AND CHAVE

Webb Data		Thurstone and Chave vs. Webb		Webb Data: Reliabilities I vs. II	
Interval	Sample	M_1	S_1	M_1	S_1
11	II_{100}	.9665	.514		
	I_{100}	.9654	.501	.9956	.841
7	II_{100}	.9524	.343		
	I_{100}	.9581	.408	.9950	.807
5	II_{100}	.9690	.413		
	I_{100}	.9643	.369	.9906	.710

S_1 values obtained on the basis of samples of 100 judges for the 11-, 7-, and 5-interval data of this study (column 4 of Table 10). (c) M_1 values of sample I_{100} vs. M_1 values of sample II_{100} for the 11-, 7-, and 5-interval data (column 5 of Table 10). (d) S_1 values of sample I_{100} vs. S_1 values of sample II_{100} for the 11-, 7-, and 5-interval data (column 6 of Table 10).

Except for extreme items the computational formula for the scale values for Thurstone and Chave and for this study would be the M_1 of this study. For ambiguity values except for extreme items, the computational formulae differ by a constant .5, since Thurstone and Chave used $(P_{75} - P_{25})$, while this study used $.5(P_{75} - P_{25})$.

A. Results

While the correlations of the Thurstone and Chave scale values with the scale values of this study are all high (approximately from .95 to .97), they are lower than the correlations between the M_1 values of the two samples for the 11-, 7-, and 5-interval data. These range approximately from .991 to .996.

The differences between (a) the correlations of the Thurstone and Chave scale values with scale values computed by the

method of this study and (b) the correlation between the two sets of scale values computed by the method of this study, were all significant at a 1% level of confidence.

The correlations between the Thurstone and Chave ambiguity values and the S_1 values of this study range from approximately .34 to .51. These are all lower than the correlations between the S_1 values of the two samples for the 11-, 7-, and 5-interval data. These range from .71 to .84. The difference between the z transformations of the correlations of the Thurstone and Chave ambiguity values with ambiguity values computed by the method of this study, and between the z transformations of the correlation between the two sets of values computed by the method of this study, were all significant at a 1% level of confidence.

B. Discussion and Summary

The results show a significant difference between ordering of items obtained by Thurstone and Chave and the ordering obtained in this study in respect to both scale and ambiguity values. When the coefficients are converted to Fisher's z, the differences in regard to scale value appear, in general, to be larger than the differences in regard to ambiguity values.

Using computational procedures which were apparently the same as those of this study, Edwards and Kenny (2) rescaled the items of Thurstone and Chave. The scale and ambiguity values of their scaling correlated .95 and .18 with the scale and ambiguity values, respectively, of Thurstone and Chave. Ten years after the construction of the Thurstone-Peterson Scale of Attitude toward War, Farnsworth (5), using the same procedures as the original authors, rescaled the 20 items of Form A. While the differences in scale values of the two scalings were significant at a 5% level of confidence for 15 of the 20 items, the

ordering of items for the two scalings gave a rank-difference correlation of .988.

While it is not possible to determine whether the differences of this study are caused by a change in computational procedure or by a change of the cultural milieu in terms of which the items are evaluated, the results of this study plus those of these other investigators suggest that the change in computational procedure has some part in the changing of the order. This suggestion follows from the fact that a lower correlation was obtained between the scale values of the two scalings when, as in this study and in that of Edwards and Kenny, both computational procedure and cultural milieu were changed, than was obtained when, as in the experiment of Farnsworth, only cultural milieu was changed. Even this suggestion must be viewed with caution, since the differences between the findings of Farnsworth and those of this study and of Edwards and Kenny may possibly be a result of differing numbers of items or differing item content.

Additional evidence in support of the effect of computational procedure on order is provided by the data regarding the relation between scale and ambiguity values discussed in the preceding section.

VIII. RELEVANCE OF SCALE AND AMBIGUITY VALUES FOR CONSTRUCTING GUTTMAN-TYPE SCALES

Edwards and Kilpatrick (3, 4) have proposed the scale-discrimination technique for selecting a set of attitude items which will have a high probability of having satisfactory reproducibility when tested by the Guttman technique. The procedure involves both the Thurstone and Likert techniques of scaling and essentially consists of eliminating half of the initial set of items with highest ambiguity and of choosing from the remaining items the desired number of items in each scale-value

interval which have the highest discrimination values.

The rationale of the technique is based on the logically derived generalizations that scale values of the Thurstone scaling process and discrimination values of the Likert scaling process are respectively related to the cutting points and reproducibility of an item in the Guttman-type method of scale analysis.

Since these generalizations were verified on small groups of items which had previously been selected by the Thurstone or Likert scaling technique, and since the efficacy of the technique was later verified by application to a basic set of items which contained very few items in the neutral range, only meager evidence of the extent or magnitude of these relationships is provided.

The purpose of this study therefore was to investigate further the interrelations among the various indices involved in the scaling process—namely, scale values, ambiguity values, discrimination values, cutting points, and item reproducibility. Results of such an investigation should be helpful in assessing the relative importance of the Thurstone- and Likert-type values for selecting items which would be scalable according to the Guttman criteria.

In order to obtain results of maximum utility, it seemed desirable that the procedures employed should meet the following requirements: (a) provide quantitative results, (b) include in the analysis all of the initial set of items and not merely those few which would be selected for scaling, and (c) include enough neutral items to prevent gaps in the distributions and thus make for greater reliability in determining curvilinearity or linearity of regression lines.

Conditions a and b require that for each of the involved items there be an index of each value expressed in quantifiable form. Such values are readily available for scale,

ambiguity, and discrimination values. Scale and ambiguity values are obtained by the Thurstone technique, while the ϕ coefficients which discriminate between the $X\%$ highest and $X\%$ lowest on total score vs. a suitable dichotomization of item responses serve as a discrimination index. But indices of cutting points and item reproducibility are arrived at only after a subset of items has been selected and even then by a method of successive approximation. It was therefore necessary to explore the possibility of finding a reasonable index of cutting point and reproducibility which could be obtained without employing the conventional Guttman-type analysis.

Though an index for item reproducibility which would satisfactorily meet these requirements could not be devised, a usable index of cutting point which met these requirements seemed available. Consider the fictitious data of Table 11 which shows a display of responses for one item tabu-

TABLE 11

DISPLAY OF FICTITIOUS DATA FOR A SINGLE ITEM TABULATED BY THE CORNELL TECHNIQUE

Rank Order of Subjects' Scores	Item Response Categories				
	Agree 1	2	Undecided 3	Disagree 4	5
20	X				
20		X			
19	X				
19		X			
18	X				
18			X		
17		X			
17			X		
17		X			
16		X			
16				X	
15			X		
15				X	
14					X
14			X		
14				X	
13					X
13					X
12				X	
12					X

lated according to the Cornell technique for performing scale analysis. The left-hand column shows the rank order of subjects according to score; opposite each score an "X" indicates the subject's response to the item. The cutting point refers to "that place in the rank order of subjects where the most common response shifts from one category to the next" (3, p. 103). In this example cutting points could be established at the appropriate points which mark the shift in predominant response from 1 to 2, 2 to 3, 3 to 4, and 4 to 5. But if, as a means of reducing error, categories for which responses intermingle are reduced to the extent that only two response categories remain, there would be only one cutting point. For this example the cutting point would fall between scores 17 and 16. The most appropriate place for establishing the dichotomy would be between response categories 2 and 3. Now since the decisions of where to establish the cutting point and the dichotomization point are made jointly on the basis of the patterning of responses, and since the determination of the point of dichotomization determines the marginal total or percentage of responses falling in each of the two response categories, it follows that the cutting point and the marginal total are related. Since the marginal total would include responses which, from the standpoint of reproducibility, constitute errors, the relation would not be perfect. But since minimization of error is one criterion considered in establishing a cutting point, the effect of this disturbance on the marginal total as an index of cutting point should not be large.

A proper determination of the cutting point and point of dichotomization can properly be made only from an array of data such as is shown in Table 11. But in view of the relations described above, it appears possible that an estimate of ac-

ceptable points of dichotomization could be made on the basis of an inspection of the distribution of responses used for computing discrimination values.

Since, as a general rule, the point of dichotomization employed in the computation of the discrimination values would probably be made so as to equalize the percentage of responses in each response category for each item, this dichotomization point should serve as a satisfactory one for establishing marginal totals as well. In this study, therefore, the percentage of response on one side of the point of dichotomization used in computing discrimination values has been used as an index of cutting point.

Since a suitable index of item reproducibility was not devised, the study has been limited to an investigation of the interrelationships among scale values, ambiguity values, discrimination values, and cutting-point indices.

A. Procedure

Indices of scale and ambiguity values used were the M_1 and S_4 values computed on the basis of sample 11-II-100. The computation of these has been described in section III.

Discrimination values and cutting-point indices were computed on the basis of responses collected by the Likert technique from 304 students enrolled in sociology and psychology classes at the Atlanta Division of the University of Georgia.

Items were ordered according to the Thurstone and Chave scale values; all items having scale values below 5.3 were judged to be favorable items, and all items above that point were judged to be unfavorable items. Scoring weights ranging from 0 for a response of strongly agree to 4 for a response of strongly disagree were assigned to favorable items; reversed weights were assigned unfavorable items. Each subject's responses were scored with these

weights. The scores ranged from 394 to 75, with a median of 164. Students were divided into upper and lower half; and for each half the frequency of response in each response category was tabulated.

Next, for the total group of subjects the distributions of responses in the five response categories for each item were considered in order to determine where to place the point of dichotomization. Inspection of the data suggested it should be between 1 and 2 or between 2 and 3. For each item the percentage of subjects responding 0 and 1 and 0, 1 and 2 was computed. These two sets of percentages correlated .966; but since the percentages based on responses 0, 1, and 2 gave a more evenly spread distribution, the point of dichotomization was chosen to fall between 2 and 3. The percentages of responses falling in the 0, 1, and 2 categories were considered as the marginal totals or cutting-point indices. After dividing the total group into upper and lower halves according to score and at the point of dichotomization of item response, ϕ coefficients were computed by the use of Jurgensen's (6) table. These were considered as discrimination values.

Finally, selected correlations among the scale, ambiguity, cutting-point, and discrimination values were determined.

Since most of the relationships were curvilinear, it was necessary to decide which variable should be considered independent and which dependent; and since discrimination and cutting-point values were the best available indices of item reproducibility and cutting points, respectively, unbiased correlation ratios have been computed for the regression of these values on the other indices. The results are shown in Table 12. For each relation the regression line used was that of the first-named value on the second.

A good deal of the curvilinearity in the scattergrams for discrimination value vs.

TABLE 12

SELECTED UNBIASED CORRELATION RATIOS BETWEEN SCALE, AMBIGUITY, CUTTING-POINT, AND DISCRIMINATION VALUES

Variables	Not Reflected				Reflected			
	K	ϵ	ϵ^2	Shape of Regression Equation	K	ϵ	ϵ^2	Shape of Regression Equation
Discrimination value vs. cutting point	10	.079*	.281	Curvilinear	10	.181†	.428	Curvilinear
Discrimination value vs. scale value	11	.338†	.571	Curvilinear	11	.192†	.438	Curvilinear
Discrimination value vs. ambiguity value	12	.204†	.442	Curvilinear	12	.048	.219	Linear
Cutting point vs. scale value	11	.810†	.000	Linear				
Cutting point vs. ambiguity value	12	−.019	−.138	Linear				
Ambiguity value vs. cutting point	10	.198†	.445	Curvilinear				
Ambiguity value vs. scale value	11	.601†	.775	Curvilinear				

* Significant at 5% level of confidence.
† Significant at 1% level of confidence.

the other indices seemed to result from the fact that the discrimination values for 12 items were negative. For these items the scoring weights were reversed so as to give positive correlations. The ϕ's were recomputed, scattergrams were replotted, and ϵ recomputed. The values so obtained are recorded under the label of "reflected values." These should be considered in evaluating the relation of discrimination values to the other variables.

B. Results

It will be observed that all of the reported relationships except three are significant at a 5% or higher level of confidence and that the regression lines for all but three depart significantly from linearity at a 1% level of confidence.

Considering the "not reflected" data the scattergram shapes of discrimination value vs. scale value and ambiguity value, respectively, were V and 7 shaped respectively. These types of shapes seemed to result from the fact that all negative discrimination values were for items in the middle range of scale values and consequently for items of high ambiguity. The regression line of discrimination value vs. cutting points was U shaped. However, when the scoring of items with negative discrimination values was reflected so as to produce positive discrimination values, the regression of discrimination value vs. ambiguity did not

depart significantly from linearity at a 5% level of confidence. The regression on scale value, however, remains roughly U shaped with lowest ϕ's falling in the scale value range of 4.0–4.9.

The regression line for discrimination values vs. cutting-point index also has something of a U shape, though since the items in the range 0.0–0.9 for cutting points had the lowest average ϕ, the curve actually has 2 flex points. Items of highest ϕ values were in the range 1.0–5.0 and 8.0–9.0.

The regression lines of cutting point vs. scale value and ambiguity value do not depart significantly from linearity. The relation between cutting point on ambiguity value was essentially zero, but the relation with scale value was relatively high. The obtained Pearson product-moment r was −.888. These data suggest no relation of cutting point on ambiguity, but since, as the data show there is a significant curvilinear regression of ambiguity value on cutting point, the ϵ of −.138[5] for cutting point on ambiguity value does not provide a full account of the relationship between these two variables. Actually, the regression line of ambiguity value on cutting point is U shaped with the U lying on its side with the open end on the left. This

[5] The negative sign results from the fact that in the process of correcting ϵ for bias, the amount of correction was larger than the obtained ϵ.

means that given a low ambiguity value one can predict a high or low cutting-point index, and given a high ambiguity one can predict a cutting point of middle range.

As previously reported, the regression of ambiguity on scale value was inverted U shaped, and the correlation ratio was significant at a 1% level of confidence.

C. Discussion and Summary

The results of this study demonstrate the following relationships: (a) an insignificant regression of discrimination and cutting-point values on ambiguity values, (b) a significant but fairly low curvilinear correlation of discrimination values on cutting-point and on scale values, (c) a moderately high curvilinear correlation for ambiguity on scale values, and (d) a high linear correlation between cutting points and scale values.

The low relation of ambiguity value to cutting-point and discrimination values together with the high relation of ambiguity to scale values suggests that knowledge of ambiguity values adds little over and above what is provided by scale value to the selection of a scalable set of items.

Further, the high relation of scale values to cutting-point values suggests that the latter could be used as a satisfactory index of ordering of items and thus obviate the necessity of obtaining scale values at all.

The fact that the discrimination values have a fairly low correlation with cutting-point and scale values suggests that this index measures something different from the other two and should be retained in the process of selecting items.

On the basis of these statements, it seems possible to infer, as Edwards and Kilpatrick (4) have suggested, that items may satisfactorily be selected on the basis of data collected by the Likert technique alone. This would mean that the pro-

cedures of the scale discrimination technique recommended by Edwards and Kilpatrick could be reduced to the computation of cutting-point and discrimination values from data collected by the Likert method, plotting discrimination vs. cutting-point values, and the selection of the desired number of items with highest discrimination values within each cutting-point index interval.

An inspection of the scattergram of scale value vs. cutting-point value suggests the possible difficulty of determining which items, if only cutting-point value were known, would fall in the neutral range in the Thurstone-technique sense. However with a set of items edited so as to include one or two items containing such specific determiners as "neutral," "indifferent to," or "don't care either way" to locate the neutral point in the neutral range, this difficulty should not arise.

It should be pointed out that the findings reported in relation to the relatively small contribution of ambiguity value over and above that of scale values is in terms of values computed by the procedures of this study. It seems probable, however, in view of the results of Edwards and Kilpatrick, that essentially the same results would have been obtained had the computational procedure of Thurstone been used.

IX. SUMMARY

This monograph reports the results of a series of studies concerning scale and ambiguity values obtained by the method of equal-appearing intervals. These studies were primarily concerned with the following topics: (1) the effects of number of intervals, number of judges, and efficiency of the estimating formulae on (a) the reliability of scale and ambiguity values, (b) the intercorrelations for scale and for ambiguity values, and (c) the curvilinear relations between scale and ambiguity

values; (2) a comparison of scale and am-
biguity values computed by the method of
this study with values obtained by the
Thurstone and Chave method; and (3) the
relevance of scale and ambiguity values for
selecting items for Guttman-type scales.
The 130 items of attitude toward the
church first employed by Thurstone and
Chave were used in the study.

For the first-named topic the procedure
involved the computation of scale and am-
biguity values from various interval-judge-
formula combinations, and analyzing the
effects of these variables on the three
aspects of scale and ambiguity values under
consideration. Three different numbers of
scale intervals (11, 7, and 5); three differ-
ent numbers of judges (100, 50, and 25);
four different formulae for computing scale
values (the mean and three estimates of the
mean based on percentile values); and four
different formulae for computing ambi-
guity values (the standard deviation and
three estimates of the standard deviation
based on percentiles) were employed. Per-
centile values for use in the computing
formulae were obtained from ogives by
dropping perpendiculars from the specified
points. This procedure differed from that of
Thurstone and Chave in the treatment of
extreme items since for these items these
investigators extrapolated for P_{50} and
doubled the quartile deviations for esti-
mating ambiguity values.

For the second topic, the ordering of
scale and ambiguity values obtained by the
procedures of this study were compared
with the ordering of items obtained by
Thurstone and Chave.

The procedure for the third topic in-
volved an analysis of interrelationships of
scale and ambiguity values computed by
the method of equal-appearing intervals,
and discrimination values and cutting-
point indices obtained from data collected
by the method of summated ratings.

The major findings of these studies were
as follows:

For scale values, reliability increases
with successively more efficient computing
formulae.

While there was a tendency for ambi-
guity values computed from increasingly
efficient computing formulae to be more re-
liable, the reliability of the standard devia-
tion was less than that of the most efficient
percentile computing formula used. In two
out of three cases it was less than that of the
least efficient formula used.

The results of a three-dimensional fac-
torial design of Fisher's z transformation of
reliability coefficients involving three num-
bers of intervals, three numbers of judges,
and three percentile computing formulae,
indicated that for scale values the judges-
by-intervals interaction and the three main
effects were significant sources of variance.
Tests of significance among the various
levels for the three main effects showed sig-
nificantly increasing reliability with in-
creasing numbers of intervals, with in-
creasing numbers of judges, and with in-
creasingly efficient formulae. However
since all reliability coefficients of all judges,
intervals, formulae were .976 or higher,
these differences are of little practical sig-
nificance.

The results from similar analyses of am-
biguity values showed that all second-order
interactions were significant at a .001 level
of confidence. There was, in addition, evi-
dence to suggest that the three main effects
were significant sources of variance. While
all differences among the various levels for
the main effects were in the expected direc-
tions, when tested against the appropriate
first-order interaction as error, they were
not all significant.

Intercorrelations computed among scale
values for selected judge-interval-formula
combinations indicated that the order-
ings of scale values computed on all the

various judge-interval-formula combinations were for practical purposes approximately the same, since all intercorrelations were .98 or higher.

Intercorrelations of ambiguity values for selected judge-interval-formula combinations showed considerable variability and suggested the advisability of determining how much departure from the ordering of items obtained from a preferred interval-judge-formula combination one will accept before using values computed on the basis of some other combination.

An inspectional analysis of the unbiased correlational ratios of ambiguity values on scale values, computed from data for selected interval-judge-formula combinations, suggested that the size of the relation for S_1 ambiguity values on M_1 scale values is a function of the number of sortings on which the values are based. For values computed on the basis of sortings of 100 judges the size of ϵ does not appear to be a function of the computing formula; but the size of ϵ for S_3 values on M_1 scale values appears to vary directly with the number of intervals.

The ordering of scale and ambiguity values obtained by the methods employed in this study differ significantly from the ordering of scale and ambiguity values obtained by Thurstone and Chave. However, it was not possible to tell conclusively whether the differences were a result of differences in procedure or differences in cultural setting within which the judgments were made.

Because of the high relation of ambiguity value on scale value, the high relation of cutting-point index on scale value, and low correlation of cutting-point on ambiguity value, it appears that cutting-point indices can be substituted for scale values, and thereby eliminate the need for scale and ambiguity value in selecting items for Guttman-type scales.

REFERENCES

1. EDWARDS, A. L. A critique of "neutral" items in attitude scales constructed by the method of equal-appearing intervals. *Psychol. Rev.*, 1946, **53**, 159–169.
2. EDWARDS, A. L., & KENNEY, K. C. A comparison of the Thurstone and Likert techniques of attitude scale construction. *J. appl. Psychol.*, 1946, **30**, 72–83.
3. EDWARDS, A. L., & KILPATRICK, F. P. Scale analysis and the measurement of social attitudes. *Psychometrika*, 1948, **13**, 99–114.
4. EDWARDS, A. L., & KILPATRICK, F. P. A technique for the construction of attitude scales. *J. appl. Psychol.*, 1948, **32**, 374–384.
5. FARNSWORTH, P. R. Shifts in the value of opinion items. *J. Psychol.* 1943, **16**, 125–128.
6. JURGENSEN, C. E. Table for determining phi coefficients. *Psychometrika*, 1947, **12**, 17–29.
7. KELLEY, T. L. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univer. Press, 1947.
8. MOSTELLER, F. On some useful inefficient statistics. *Ann. math. Statist.*, 1946, **17**, 377–408.
9. THURSTONE, L. L. Attitudes can be measured.

Amer. J. Sociol., 1928, **33**, 529–554.
10. THURSTONE, L. L., & CHAVE, E. J. *Measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.
11. WEBB, S. C. A generalized scale for measuring interest in natural science subjects. *Educ. psychol. Measmt*, 1951, **11**, 456–469.
12. WEBB, S. C. Scaling of attitudes by the method of equal-appearing intervals: a review. *J. soc. Psychol.*, in press.
13. WEBB, S. C. Irregularities in judgment data collected by the method of equal-appearing intervals. *J. abnorm. soc. Psychol.*, 1954, **49**, 415–418.
14. WICKER, M. P. A comparison of attitude scale values yielded by scales of differing lengths. Unpublished master's thesis, Univer. of North Carolina, 1950.
15. YOST, E. K. Joint estimation of mean and standard deviation by percentiles. Unpublished master's thesis, Univer. of Oregon, 1948.